



# Collection and Elicitation of Business Process Compliance Patterns with Focus on Data Aspects

Thomas Voglhofer · Stefanie Rinderle-Ma

Received: 16 February 2018 / Accepted: 4 February 2019 / Published online: 29 March 2019  
© The Author(s) 2019

**Abstract** Business process compliance is one of the prevalent challenges for companies. Despite an abundance of research proposals, companies still struggle with manual compliance checks and the understanding of compliance violations in the light of missing root-cause explanations. Moreover, approaches have merely focused on the control flow perspective in compliance checking, neglecting other aspects such as the data perspective. This paper aims at analyzing the gap between existing academic work and compliance demands from practice with a focus on the data aspects. The latter emerges from a small set of regulatory documents from different domains. Patterns are assumed as the right level of abstraction for compliance specification due to their independence of (technical) implementation in (process-aware) information systems, potential for reuse, and understandability. A systematic literature review collects and assesses existing compliance patterns. A first analysis of ten regulatory documents from different domains specifically reveals data-oriented compliance constraints that are not yet reflected by existing compliance patterns. Accordingly, data-related compliance patterns are specified.

**Keywords** Business process compliance · Compliance (anti) patterns · Data perspective of business processes · Regulatory data constraints

## 1 Introduction

Organizations in many industrial sectors are required to manage their business processes in accordance with regulations, laws, controls, and other industrial specific constraints, referred to as Business process compliance (BPC) (Rinderle-Ma and Kabicher-Fuchs 2016; Schleicher et al. 2010). The financial sector, for example, faces far more regulatory constraints since the financial crisis of 2008 that revealed the fragility of the financial system (Awad et al. 2015). Another example emphasizing the importance of data in compliance considerations is the European General Data Protection Regulation (GDPR) which became effective in 2018. Although IT support for BPC has been increased, still many organizations have to invest massively in BPC (cf. Awad et al. 2015), mainly caused by manual compliance checks (Ly et al. 2015; Rinderle-Ma and Kabicher-Fuchs 2016).

### 1.1 Motivation

Figure 1 shows the six phases of a BPC life cycle based on the phases of compliance as set out in Awad (2010). It starts with the *Regulation identification* as first phase where regulations, laws, standards, and guidelines are scanned to spot the ones relevant for the respective purpose. In the second phase *Constraint elicitation*, constraints are extracted from the relevant regulatory documents. Not knowing which constraints an organization has to comply with, might lead to violations of those constraints.

---

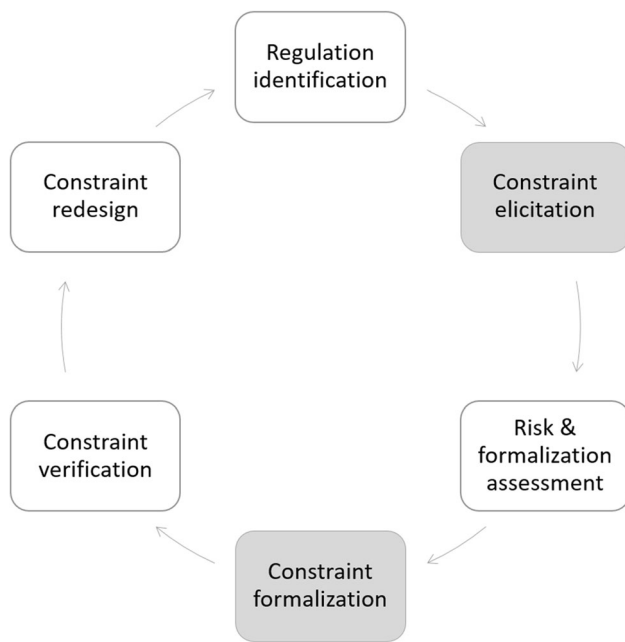
Accepted after three revisions by Jörg Becker.

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s12599-019-00594-3>) contains supplementary material, which is available to authorized users.

---

T. Voglhofer (✉) · S. Rinderle-Ma  
Research Group Workflow Systems and Technology, Faculty of  
Computer Science, University of Vienna, Vienna, Austria  
e-mail: thomas.voglhofer@univie.ac.at

S. Rinderle-Ma  
e-mail: stefanie.rinderle-ma@univie.ac.at



**Fig. 1** BPC life cycle

Consequences can be fines, economic disadvantages, and reputation damage. *Risk & formalization assessment* represents the third phase in which the determined constraints are evaluated in accordance with their impact on an organization in case of violation. Furthermore, the possibility of formalization is analyzed for each constraint and a decision is made which constraints are formalized. In the fourth phase of *Constraint formalization* the selected constraints are formalized using different formal approaches [e.g., Linear temporal logic (LTL), Event calculus (EC), Complex event processing (CEP)]. The fifth phase is *Constraint verification* which checks if the regulatory constraints imposed on business processes are fulfilled. Many approaches exist for compliance verification (e.g., Awad et al. 2015; Elgammal et al. 2010; Hashmi et al. 2012; Ly 2016; Ly et al. 2015; Turetken et al. 2011), and can be categorized into design-time, runtime, and ex-post checking of BPC according to Fellmann and Zasada (2014). In the sixth phase of *Constraint redesign*, existing regulatory constraints have to be adapted due to new or changed laws, standards, or guidelines. For this, regulatory constraints must be continually reviewed and updated. This paper focuses on the phases of constraint elicitation and constraint formalization, which are highlighted in Fig. 1. Constraint elicitation builds the basis for constraint formalization and identifies relevant constraints within regulatory documents. Constraint formalization also means to specify identified constraints in a suitable way that many people in an organization will understand. For this purpose, various approaches propose patterns in general and Compliance patterns (CPs) in particular. For instance, Dwyer

et al. (1998) and Turetken et al. (2011) advocate CPs, because complex logic and formal specifications are hidden from business users. Thus, business users become able to understand the meaning behind and make use of the CPs. This property of CPs also fosters the common understanding of a certain, recurring problem/issue between different departments within an organization, for instance between the legal unit and IT which has to implement the defined compliance measures. CPs are independent of approach and language which means that they are universally applicable to and (re)usable for various regulatory constraints. So no extra effort is necessary when changing between different languages and approaches. They also facilitate a solid basis for the analysis of root-causes in case of compliance violations.

## 1.2 Problem Statement

CPs offer the right level of abstraction. For furthering BPC support in practice a comprehensive collection of CPs that is based on real-world needs would be crucial. Comprehensive in this context refers to the coverage of typical business process perspectives that CPs might relate to, i.e., control flow (including order and occurrence), data, resources, and time (Ly et al. 2015). Most approaches suggest and use control flow related CPs. Becker et al. (2012) compare different approaches by positioning compliance approaches into the support of CPs of simple, medium, and high complexity where only 2 approaches support highly complex CPs referring to data or time. For instance, precedence and consequence patterns for activities in a business process are used several times as examples (e.g., Awad et al. 2015; Barnawi et al. 2016; Chesani et al. 2008; El Gammal 2012; Yu et al. 2006). The focus of BPC publications on the order and occurrence perspective is also highlighted by Fellmann and Zasada (2014). Some papers also include a small collection of up to 15 CPs (Awad 2010; De Masellis et al. 2014; Namiri and Stojanovic 2007). However, a broader collection and a deeper analysis of existing CPs like done by Caron et al. (2013a) and Ramezani (2017) are rare, although they support the assumption that also process perspectives beyond control flow are crucial for CPs. In addition, only few approaches deal with CPs in the context of more than one domain (Awad 2010; Bernardi et al. 2014; El Gammal 2012; Elgammal et al. 2016; Gong et al. 2016; Ly 2016; Ly et al. 2015; Ramezani 2017), but often lack the elicitation of CPs from those domains. Thus, an investigation of a broad range of regulatory documents is missing.

Overall, this discussion emphasizes the need for a systematic and comprehensive CP collection. This paper takes up this challenge by addressing the following research questions (Qs):

- Q1: Is there a gap between coverage of business process perspectives in literature and demands from regulatory documents? Particularly for data-oriented compliance demands?
- Q2: Which data-oriented compliance constraints in regulatory documents are not covered by existing CPs?
- Q3: How can missing CPs be defined for uncovered data-oriented compliance constraints?

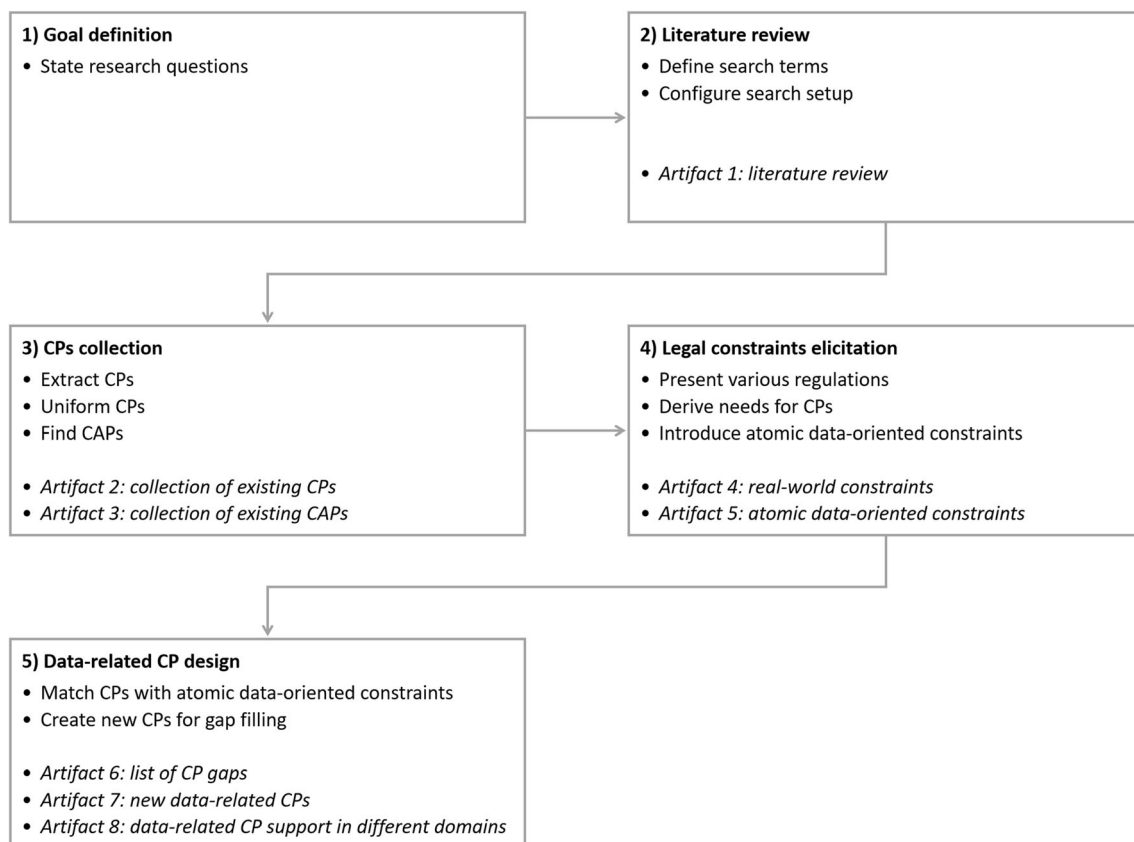
Q1 aims to assess the comprehensiveness of existing CP collections. Q2 relates to gaps in the data perspective of business processes regarding coverage of regulatory constraints. In case that CPs which are required from real-world applications are missing in current collections, Q3 aims at filling this gap.

### 1.3 Method and Contribution

Figure 2 depicts the overall research method employed in this paper including the artifacts created throughout the applied method. Starting point is the goal definition based on the research questions. The literature review performed in the second step identifies existing CPs in research and the current status of research concerning the topic of CPs in general ( $\mapsto$  *Artifact 1*). All identified CPs are extracted into

a uniformed collection of CPs ( $\mapsto$  *Artifact 2*) as proposed by Caron et al. (2013a). The same procedure is applied to compliance anti-patterns (CAPs) ( $\mapsto$  *Artifact 3*), which are in most cases the negation of CPs and can be seen as a subset of CPs. In addition, if possible the relation between CP and CAP is established. The legal constraints elicitation step identifies compliance constraints stated in various regulatory documents ( $\mapsto$  *Artifact 4*). Those legal and regulatory documents are selected due to their generality of application area, their different domains, and their up-to-dateness. The compliance constraints serve as basis to state atomic data-oriented constraints ( $\mapsto$  *Artifact 5*). The fifth step compares existing CPs to elicited atomic constraints previously identified in Step 3. If constraints cannot be mapped to CPs, they are collected in a separate list of gaps ( $\mapsto$  *Artifact 6*). The atomic constraints are then further used to derive data-related CPs ( $\mapsto$  *Artifact 7*). Special focus is on data-related CPs to emphasize the increased importance of data for economic, competitive, and legal reasons. Furthermore, the support of CPs in different domains is analyzed ( $\mapsto$  *Artifact 8*).

The provided results are supposed to support business users in selecting the CP of interest and need. The results additionally enable users to view the entire set of CPs from



**Fig. 2** Overall method

various perspectives. The intention is to further the research concerning the root-cause analysis of compliance violations in business processes.

The paper is structured as follows. In Sect. 2 the criteria for and the results of the performed literature review are listed. Section 3 analyzes different legal and regulatory documents that require compliance for data. In Sect. 4 a comparison between current constraint needs and existing CPs as well as the design of new data-related CPs is conducted. The advantages and disadvantages of the applied method and the results of the paper are discussed in Sect. 5. The paper concludes with a summary and an outlook in Sect. 6.

## 2 Literature Review

Compliance assurance, compliance violation detection, compliance validations, and the compliance topic for business processes in general form a broad area of research. Thus, several terms are used for similar and/or the same concepts. A literature review helps to get an overview of the current status quo of research concerning compliance in business processes with special focus on the use of CPs for compliance monitoring, compliance violation and root-cause identification, as well as compliance verification and validation.

The literature review follows and adopts the guidelines stated by Kitchenham and Charters (2007). This section describes first the process performed to conduct the literature review (Sect. 2.1) and the content of the retrieved literature (Sect. 2.2). The second part deals with the results of the literature review (Sect. 2.3) and the CP collection (Sect. 2.4). Section 2.5 shortly summarizes the literature review. Altogether this section builds the core of *Artifact 1: literature review*.

### 2.1 Search Process

The search process starts with the definition of inclusion and exclusion criteria for literature. Then search strategy and data extraction are determined. Afterwards the literature is synthesized and prepared for reporting.

The inclusion criteria for literature selection are:

- *Accessibility* – The literature must be freely accessible. That includes all literature available via Internet search.
- *Language* – The literature must be written only in English, which makes the retrieved results comparable and helps to minimize the ambiguity of key terms like pattern and constraint.
- *Journal, author, and date of publication* – No limitations are set on the journal, author, or date of

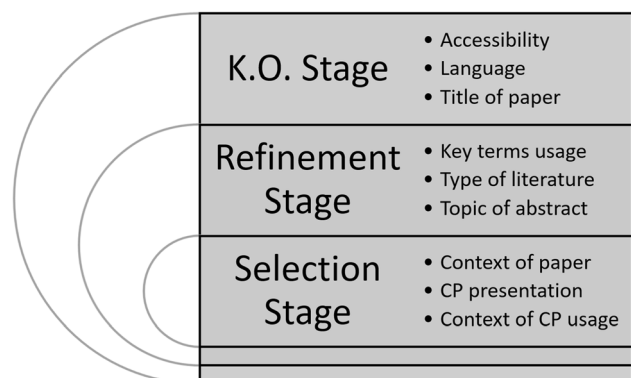
publication. That increases the possibility to find as much relevant literature as possible and builds a comprehensive overview of available CPs in literature.

- *Title of paper* – The title must include the search terms in any order. That fact highlights the importance of those terms in literature and helps to retrieve literature using different orderings of terms in the title.
- *Key terms usage* – The terms *pattern(s)*, *rule(s)* and *constraint(s)* must be used in the text of the retrieved literature.
- *Context of paper* – Those terms listed by inclusion criterion *Key terms usage* have to be used in the context of business processes, CPs, compliance frameworks, or compliance constraints.
- *CP presentation* – The literature must describe (certain) CPs in a way that the meaning of the CP is clear and/or a graphical representation of the CP is available.

The exclusion criteria negate the inclusion criteria, but are extended with further criteria:

- *Type of literature* – The literature review excludes literature that is not published via a digital library, journal, or in conference proceedings. Bachelor and Master theses as well as books are excluded from further processing.
- *Topic of abstract* – Documents with an abstract that does not deal with CPs or directly related issues are excluded.
- *Context of CP usage* – Literature using patterns in context of software engineering/development, social sciences and other domains than specified in the inclusion criterion *Context of paper* are excluded.

The search strategy is split into three stages with combined usage of inclusion/exclusion criteria (cf. Fig. 3). In the first stage the K.O. criteria *Accessibility*, *Language*, and *Title of paper* exclude non-relevant literature. Those criteria are the entrance criteria for further investigation of the papers. In the next stage the filtered literature is further analyzed



**Fig. 3** Stages of search strategy

based on first content-related criteria to refine the search results. In the selection stage a detailed review of the remaining literature is performed to get the final literature list.

The search strategy makes use of a search engine and applies a horizontal as well as a vertical search. In this context the horizontal search tries to retrieve as many search hits as possible for the given search strings, whereas the vertical search exploits the search hits' references of the horizontal search as well as experts' advices to retrieve specific details. The search engine must be online and freely accessible with any device, and must allow as well to search for terms in the title of literature. In addition, the search engine must allow to concatenate search terms with Boolean operators and to search for literature in English only.

Google Scholar is used as search engine, because it fulfils these criteria. Besides that, Google Scholar is not limited to a single journal, conference proceedings, or dedicated digital libraries. In addition, *u:search* – the search engine of the library of the University of Vienna – has been facilitated to retrieve literature which is not freely available via Google Scholar.

The search strategy uses concatenations of search terms to search strings. Search strings allow to narrow the search space, but are also able to cover various variants of search terms (e.g., singular and plural of terms). Kitchenham and Charters (2007) suggest to divide the research questions into separate parts and list “synonyms, abbreviations, and alternative spellings” to find search terms. The following terms are determined according to research questions Q1 and Q2: *business, process, compliance, constraint/(anti)-pattern/rule, validation/verification, framework, monitoring, violation, runtime/design-time, mining, rule-based, model-based, and data*. Based on them 18 different search strings are composed for the horizontal search. These include the terms listed in Table 1 in various combinations.

In addition to the CPs their description and categorizations/classifications as well as the year of publication are extracted. The latter facilitates the analysis of the domain and publication of literature, and crucial events of recent times (e.g., the financial crisis of 2008).

Based on the criteria for literature selection and search engines a horizontal literature search is performed. For this

the search terms must be included in the title of the documents and only documents in English are searched as described in the literature search's inclusion criteria *Language* and *Title of paper*. “Appendix A” lists the detailed results of the horizontal literature search including a list of the selected literature. The three columns correspond to the three search stages. If documents are retrieved multiple times, the first occurrence of a document is listed. Each search term combination is stated in the first column grouped by the date when the search was performed. The second column represents the number of papers retrieved using the corresponding search terms. Out of 798 distinct hits in total, 111 hits are investigated in detail and 34 are finally selected based on the previously defined inclusion and exclusion criteria.

In addition, a vertical search is performed and additional literature is identified by experts (see “Appendix B”) resulting in 17 additional documents in total, out of which 13 documents are finally selected. Altogether, 34 + 13 = 47 papers form the literature list for further consideration.

## 2.2 Literature Summary

The majority of existing approaches includes CPs. Only seven papers explicitly deal with CAPs (Awad 2010; Awad and Weske 2010; Awad et al. 2011, 2015; Barnawi et al. 2016; Montali et al. 2014; Trčka et al. 2009). Lu et al. (2009) use CPs and Bernardi et al. (2014) use CAPs but they do not name them as such. The compliance monitoring/management approaches are applied to or created from different domains such as electronic business (Elgammal et al. 2010, 2016; Papazoglou 2011; Turetken et al. 2011; Yu et al. 2006) or higher education (Lam 2017; Ly 2016; Ly et al. 2015).

Around one third of the papers proposes their own compliance monitoring or compliance management framework (Awad et al. 2015; Barnawi et al. 2016; El Gammal 2012; Elgammal et al. 2016; Giblin et al. 2006; Gong et al. 2016; Ly 2016; Ly et al. 2011, 2015; Maggi et al. 2011; Montali et al. 2014; Papazoglou 2011; Schumm et al. 2010; Thullner et al. 2011; Turetken et al. 2012). Those frameworks mainly try to comprehensively address the topic of compliance verification by checking compliance at design-time (Awad 2010; Awad and Weske

**Table 1** Search terms

Business	Process	Processes	Compliance	Anti
Pattern	Patterns	Monitoring	Constraint	Constraints
Framework	Mining	Rule	Rules	Rule-based
Validation	Validations	Verification	Verifications	Violation
Violations	Runtime	Model-based	Design-time	Design time
Data				



2010; Awad et al. 2009, 2011; Becker et al. 2010; Bernardi et al. 2014; Cheikhrouhou et al. 2014; El Gammal 2012; Elgammal et al. 2010, 2016; Gomez-Lopez et al. 2013; Ly 2016; Ly et al. 2011; Schumm et al. 2010; Stuht et al. 2012; Turetken et al. 2012) or checking compliance based on execution traces at the runtime of a business process (Awad et al. 2015; Barnawi et al. 2016; Chesani et al. 2008, 2009; De Masellis et al. 2014; El Gammal 2012; Lam 2017; Ly 2016; Ly et al. 2015; Maggi et al. 2011; Montali et al. 2014; Ramezani 2017; Santos et al. 2012).

In order to identify compliance violations multiple formalisms are used. LTL is a very commonly used one (Bernardi et al. 2014; Caron et al. 2013a, b; De Masellis et al. 2014; Dwyer et al. 1998; El Gammal 2012; Elgammal et al. 2010, 2016; Gong et al. 2016; Lam 2017; Maggi et al. 2011; Schumm et al. 2010; Stuht et al. 2012; Turetken et al. 2012). Other formalisms in literature are (colored) automata (Cheikhrouhou et al. 2014; De Masellis et al. 2014; Gruhn and Laue 2005; Maggi et al. 2011; Santos et al. 2012; Gruhn and Laue 2006), Computational tree logic (CTL) (Awad and Weske 2010; Awad et al. 2011; Dwyer et al. 1998; Stuht et al. 2012), First order logic (FOL) (Caron et al. 2013a, b; Ly et al. 2010), CEP (Awad et al. 2015; Thullner et al. 2011), Logic programming (Chesani et al. 2008, 2009), (Mixed) Integer programming (IP) (Kumar and Barton 2017; Kumar et al. 2010, 2015), EC (Montali et al. 2014), Declare (van der Aalst et al. 2017), and Business process constraint network (BPCN) (Lu et al. 2009).

Some papers focus on the graphical representation of compliance constraints, CPs and violations using Compliance rule graphs (CRGs) (Gomez-Lopez et al. 2013; Knuplesch and Reichert 2017; Ly 2016; Ly et al. 2010, 2011). The graphical presentation eases the interaction with and use of CPs and the results of compliance monitoring for business users.

Few approaches go beyond the “simple” identification of compliance violations and try to offer fixing actions (Awad et al. 2009; El Gammal 2012; Elgammal et al. 2010; Maggi et al. 2011; Ramezani 2017; Lu et al. 2009). Such remedy strategies include, for instance, change of the process model structure (Awad et al. 2009), ignorance or reset of violations, change of business process execution (Maggi et al. 2011), and analysis of violation context (Ramezani 2017).

Schumm et al. (2010) address compliance from another direction by trying to ensure compliance by design. For this they propose to use predefined and already compliance ensured process pieces to integrate compliance constraints into a business process.

Lu et al. (2009) use CPs to apply constraints to process variants which are used in their framework to adapt process instances by domain experts if needed. For this they define

selection and scheduling constraints which represent occurrence and order CPs.

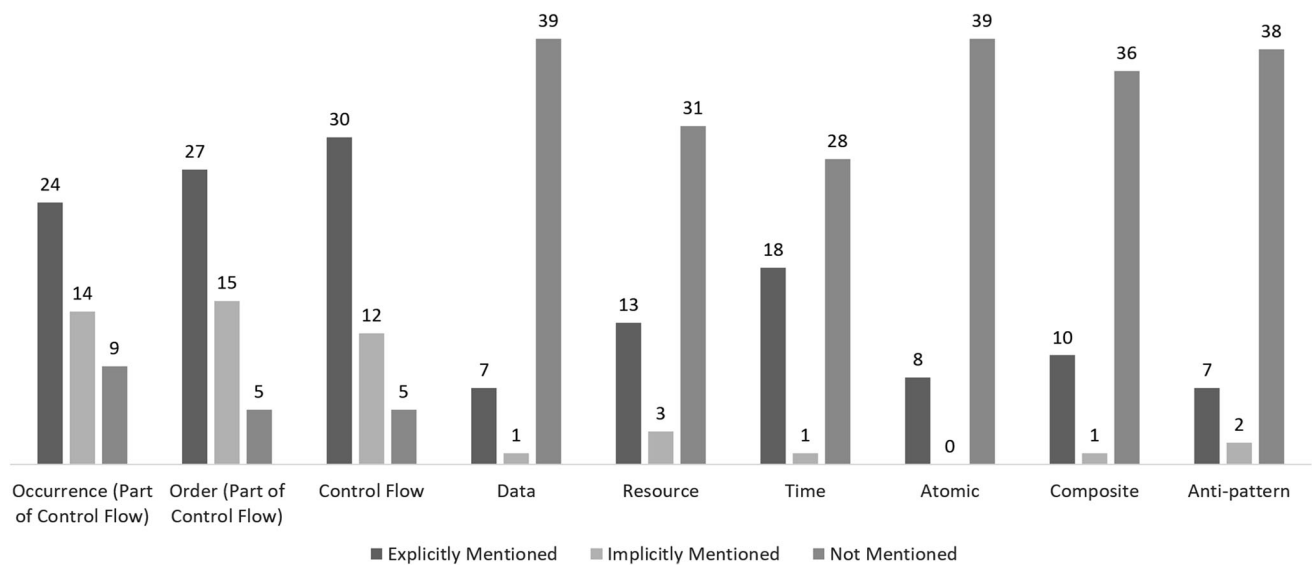
Cabanillas et al. (2010) describe what data-related compliance problems exist and how they can be grouped. Some of the mentioned problems can be represented as CPs.

### 2.3 Literature Analysis

This section analyzes compliance coverage of perspectives and domains. Doing so it contributes to *Q1: Is there a gap between coverage of business process perspectives in literature and demands from regulatory documents? Particularly for data-oriented compliance demands?*

Typical business process and compliance perspectives are control flow, resources, time, and data (Ly et al. 2015). Additional properties found in literature are atomic and composite CPs and CAPs. The atomic property relates to individual CPs, whereas CPs assigned to the composite property are two or more CPs combined with Boolean operators. Figure 4 shows how these perspectives and properties are addressed in compliance literature and a full matrix of the used literature and perspectives and properties is provided in “Appendix C”. The number of documents using a perspective or a property are represented as the bars. Three different categories of usage exist. If perspectives or properties are used by literature this is indicated by *Explicitly mentioned*. If literature does not mention any perspective or property, we assume one or more perspectives and properties based on the given CPs, which is then listed as *Implicitly mentioned*. *Not mentioned* is taken if the respective perspective or property are not included in literature in any way.

The three perspectives *occurrence*, *order*, and *control flow* basically represent the same overall category of *control flow*, but sometimes *occurrence* and *order* are listed as distinct perspectives (e.g., Awad 2010; Awad et al. 2011, 2015; Caron et al. 2013a), whereas both of them are sometimes combined under *control flow* (e.g., Becker et al. 2010). The *control flow* perspective is the most frequently mentioned perspective (30 papers). Second is the *time* perspective (18 papers) and third the *resource* perspective (13 papers). Far off from the other three business process perspectives is the *data* perspective. It is only mentioned by 7 papers, which is roughly half of the number for resource, roughly three times less than for time, and more than four times less than for the control flow perspective. Also if the numbers of papers which just implicitly mention a perspective are viewed, the data perspective has the lowest number. Thus, the data perspective shows the highest number in regards of unmentioned counts within the investigated documents, too. This finding does not reflect that Data quality (DQ) is one of the leading



**Fig. 4** Coverage of process perspectives and properties in compliance-related literature

challenges for today's organizations (Paulson 2000), since data becomes more and more a crucial commodity like water, oil, and steel and represents an existential cornerstone of today's and future (fully) automated industry as well as the foundation for every fact-based board decision (Marín-Ortega et al. 2014). Also in the research area DQ has to be ensured in order to obtain reliable results. Therefore different approaches are used to achieve this goal (Khan et al. 2012; Stausberg et al. 2011). If available, data-related CPs can be even integrated into organizations' DQ management frameworks to support overall DQ.

Less frequently, existing approaches utilize the properties *atomic*, *composite*, and *anti-pattern* where *atomic* is explicitly mentioned eight times, *composite* ten times, and *anti-pattern* seven times.

Finally, we take a look at the number of real-world domains a single document applies its approach to. Typically the research is only applied to one domain like maritime safety, online product selling application, or supply chain management (e.g., Awad and Weske 2010; Cheikhrouhou et al. 2014; Chesani et al. 2008; Elgammal et al. 2010; Ly et al. 2011; Santos et al. 2012; Yu et al. 2006). Nevertheless, about one fourth of the papers deal with two or more domains like banking and e-business, health care, manufacturing, higher education, maritime safety, and IT projects, or Internet reseller and loan origination and approval (cf. Awad 2010; Bernardi et al. 2014; El Gammal 2012; Elgammal et al. 2016; Gong et al. 2016; Ly 2016; Ly et al. 2015; Ramezani 2017; Turetken et al. 2012). Also three documents do not state any application domain at all (cf. Caron et al. 2013b; Gruhn and Laue 2005, 2006).

## 2.4 CP Collection

CPs are extracted from the final literature list to create a CP collection. The extracted CPs are then integrated into the existing CP taxonomy as proposed in Caron et al. (2013a). They propose a comprehensive rule-based compliance checking approach. This approach rests on three main architectural components: Business provenance records the actual and past business events. Legislation, policies, and other directives deal with various kinds of constraints. Techniques consist of rule patterns (i.e., CPs) and rule-based controls. A so called business rule taxonomy includes all CPs structured in accordance with the two main perspectives Process mining perspective (PMP) and Rule restriction focus (RRF) perspective. The PMP covers three out of four business process perspectives (control flow, resource, and data) and the RRF perspective includes the fourth business process perspective of time.

This taxonomy is by far the largest one we found and supports the main business process perspectives *control flow*, *resources*, *time*, and *data*. Thus, a substantial foundation for the collection of CPs exists and allows to categorize CPs more easily accordingly to the given taxonomy.

The collection process is detailed for CPs (*Artifact 2*) and CAPs (*Artifact 3*): First CPs are identified in literature. For this we look for CPs enumerations, their formal representation, and the various substitutes, e.g., semantic rules and compliance requirements. If a CP is identified, an investigation of its meaning is performed. Often only the name and a textual description are available, sometimes a formalism to get the intention behind the CP. After a clear understanding of the CP goal, a search in the existing CP collection is performed. If a corresponding CP matches, the

CP source is added as reference to the collection. Otherwise, the CP is transformed to the structure of the CP collection (cf. Caron et al. 2013a). For instance the identified CP *Exists* in Turetken et al. (2011) *describes a condition that necessities the existence of P* is mapped to *An activity of type  $a_1$  must be performed at least once* which represents the structure of CPs in Caron et al. (2013a). Afterwards the transformed CP and the CP source are added to the CP collection. This process is conducted for every CP as well as every CAP in the investigated literature.

In total, 215 CPs/CAPs are identified from the 47 papers of the final literature list. They include the 64 CPs already contained in the CP collection offered by Caron et al. (2013a), but extend this collection into the – to the best of our knowledge – most comprehensive CP collection (*Artifact 2, Artifact 3*) based on existing literature.<sup>1</sup> Moreover, the CP collection builds the basis for matching elicited compliance constraints, i.e., based on the collection it can be decided whether a compliance constraint can be assigned to an existing CP in the collection or potentially a new CP has to be defined.

## 2.5 Conclusion

The analysis of the investigated literature shows that all business process and compliance perspectives are covered (i.e., mentioned in at least one paper) by literature. Some perspectives are obviously highly preferred above others.

Another finding is that the data perspective falls short in explicit coverage and awareness when compared to the other perspectives. This perspective will need further attention in future, so all compliance constraints concerning data can be sufficiently fulfilled. Examples for regulations containing data-oriented constraints are the *Guidelines on anti-money laundering & counter-financing* as referred to in Awad (2010) and Awad et al. (2011). Ramezani (2017) uses an internal policy from the Dutch national Employee Insurance Agency. These examples will be augmented with a detailed analysis of further regulatory documents with respect to the need for supporting data-oriented constraints in the following Sect. 3.

## 3 Analysis of Regulatory Documents

Economic and competitive advantages (Paulson 2000) as well as legal and regulatory constraints (e.g., DPA 2000; Bank for International Settlements 2013) enforce a detailed understanding and consideration of data read, written, and

transformed through business processes. This section analyzes regulations stemming from different domains, including a variety of compliance constraints related to all business process perspectives and properties of CPs. As the conclusion of Sect. 2.5 is that the data perspective has been underestimated by existing compliance approaches, particular focus is set on the existence and type of data-oriented constraints within the analyzed regulatory documents.

The following method is applied: In a first step regulatory documents are searched along the following requirements: The regulatory documents are selected from different domains/industries to give a balanced overview of current and future constraints in organizations (*RegC1*). Secondly, the regulatory documents should demand for processing of data (*RegC2*). Finally, the regulatory documents must contain currently implemented and if available also future regulations (*RegC3*). The second step of the method deals with the identification and extraction of constraints from the regulatory documents. Afterwards atomic data-oriented constraints are derived from the extracted constraints. They serve as input for the CP design in Sect. 4. The following sections detail those steps.

### 3.1 Search for Regulatory Documents

The selected documents cover the domains of financial industry (i.e., AnaCredit; Bank for International Settlements 2013), health care (i.e., ELGA-VO 2015; GTelG 2012), IT security (BSI Act 2009), energy sector (IMA-VO 2011; Oesterreichs Energie 2015, 2018), data protection (i.e., DPA 2000) and e-government (i.e., E-GovG). All of them include constraints that affect the processing of data. DPA (2000); E-GovG; GTelG (2012), IMA-VO (2011), Oesterreichs Energie (2015, 2018) and ELGA-VO (2015) are already implemented regulations in Austria. The BSI Act (2009) is effective in Germany. The other two regulations AnaCredit and Bank for International Settlements (2013) must be implemented and complied to in the next one to two years by organizations in Austria (cf. “Appendix D”). Those domains are selected due to their broadness and versatility, their use in other papers, and thus their well-known level of awareness. The selection criterion *RegC3* is split into two separate columns in “Appendix D” to better differentiate between the effective dates of the regulations. All selected regulatory documents fulfil all three criteria and the details are listed in “Appendix D” where a ✓ shows the fulfilment of a criterion. The Bank for International Settlements (2013) developed 14 “Principles for effective risk data aggregation and risk reporting” – also known as BCBS 239 – which require banks to aggregate and report risk data in a complete, accurate, and timely way by using as well data taxonomies including metadata and naming conventions.

<sup>1</sup> The CP collection can be found here: <https://wst.cs.univie.ac.at/research/projects/project/8/>.



Regulation (EU) 2016/867 of the European Central Bank (ECB) on the collection of granular credit and credit risk data (ECB/2016/13) also known as AnaCredit controls the reporting of credit data to the national central banks and the ECB, respectively. The reported credit data must, for instance, follow dedicated rules concerning accuracy, and data format and type (AnaCredit).

The Data Protection Act 2000 (DPA 2000) is more general in regards to data-related compliance. It deals with the trustworthy and correct processing and storage of personal and sensitive data (i.e., data that identifies or makes data subjects identifiable and data concerning racial or ethnic origin, political opinion, and religious beliefs). All data processing has to ensure that used data are factually correct by means of reasons of application.

The ELGA-Verordnung 2015 (ELGA-VO 2015) deals with the implementation and improvement of the Electronic health record (EHR) and includes data-oriented constraints. The main data focus is on the completeness of data and the correct data delivery according to specific conformity criteria.

Another health industry-related Austrian law is the Federal Act on Data Security Measures when using personal electronic Health Data (GTelG 2012). It deals with the handling and usage of personal electronic health data to ensure a minimum set of standards for data security, extend the information basis on e-health and define rules for undirected communication of electronic health data.

The Federal Act on Provisions Facilitating Electronic Communications with Public Bodies (E-GovG) promotes the electronic legally relevant communication with public bodies. Crucial cornerstone of this law is the unique electronic identification of legal persons. Thus, it defines electronic identity, how to get an electronic identity and how it shall be used for the communication with public offices.

The Act on the Federal Office of Information Security (BSI Act – BSIG) (BSI Act 2009) describes the tasks, responsibilities, and authorization of the federal office. It “shall promote the security of information technology” through, for instance, analyzing security risks, testing and evaluating security of IT and servicing other federal bodies with security products. The office shall further protect communications technology of the German Federation and serves as central contact point for critical infrastructure operators in regards to IT security.

The Intelligente Messgeräte-AnforderungsVO (IMA-VO 2011) defines the requirements for smart metering utilities. It describes the application area and technical as well as process-related requirements (e.g., bi-directional communication possibility, storage capacity, timing requirements).

The Requirements Catalog End-to-End Security for Smart Metering (Oesterreichs Energie 2018) “describes the minimum requirements for end-to-end secured Smart Metering for electricity in Austria. These requirements apply to manufacturers during tender processes for the Smart Meter, Gateway, Central System, and their communication links.”

The Smart Metering Use-Cases (Oesterreichs Energie 2015) give an overview of uses cases that shall be supported in a smart metering system in Austria (e.g., data read out, deactivation of smart meter, prepayment). The document uses different regulatory documents as basis (e.g., IMA-VO 2011).

### 3.2 Extraction of Atomic Data-Oriented Constraints

We go through each of the regulatory documents and identify and extract atomic data-oriented constraints in a manual way. Only those constraints are considered that impose restrictions on data read, written, and transformed in business processes, i.e., constraints that are checked and enforced in the context of a business process and not elsewhere (e.g., in a database). For instance, Article 5 (1) of AnaCredit states *Credit data shall be reported (...) where the debtor’s commitment amount is equal to or larger than EUR 25, 000 on any reporting reference date within the reference period..* Here, several atomic data-oriented constraints (e.g., amount has to be equal to or larger than 25,000, amount has to be in Euro, amount has to be from respective reporting period) can be defined to ensure a correct credit data reporting to the supervisors (in the associated business process). Table 2 contains the extracted atomic data-oriented constraints on the left hand side (*Artifact 5*), e.g., *7 Data must be in domain*. Altogether 19 constraints are identified and extracted. The total number of occurrences of an atomic data-oriented constraint per regulatory document is listed in the respective columns of Table 2 to the right.

In summary, the analysis of regulatory documents shows the relevance of compliance constraints in general (*Artifact 4*): in each of the documents at least 4 different atomic data-oriented constraints were identified and extracted. AnaCredit includes 12 different atomic data-oriented constraints which is the highest number. Table 2 shows in detail which regulatory document uses which atomic data-oriented constraints. It further can be deduced that the extracted constraints refer to one or several of the four business process perspectives control flow, data, time, and resources. The scan of the regulatory documents strengthens the presumption that constraints referring to other business process perspectives than data might be covered by already existing CPs. For instance, *Principle 1* of BCBS 239 requires the active participation of the board and senior

**Table 2** Occurrence of atomic data-oriented constraints in regulatory documents

No.	Atomic data-oriented constraints	Ana credit	BCBS 239	BSI act	DSG 2000	E-GovG	ELGA-VO	GTelG	IMA-VO	Req. Cat. SM	SM UC
1	Data must be accurate/consistent/must have integrity	1	8		2					9	
2	Data must be available at/for a certain time		1	7	2	2	1		2		3
3	Data must be available/complete	4	6	2	4		4	1		20	119
4	Data must be calculated using formula	25	1			7	1	1		3	
5	Data must be equal/have the same values									1	
6	Data must be from a certain time span	3	1		1				1		5
7	Data must be in domain	37			1	1	4	5			8
8	Data must be in range	6									4
9	Data must be missing	8		4		1				1	3
10	Data must be of certain granularity	7						1			
11	Data must be of specific data type	66								1	43
12	Data must be unique	21	1		1		7	7		3	2
13	Data must be unique over time	8									
14	Data must conform to a specific format	80	3		4	1	9	7		3	9
15	Data must not be change									2	
16	Data must not be used			1							
17	Data must be encrypted/decrypted					4			1	5	4
18	Data origin must be known				2						
19	Data purpose must be known				1						
Grand total		266	21	14	18	16	26	21	5	48	200

management of a bank which can be covered by CP *performed by* (cf. Barnawi et al. 2016; El Gammal 2012; Ly et al. 2015). Another example are paragraphs 4 and 6 of Article 13 stated in AnaCredit where national central banks have to transfer monthly credit data until a defined deadline being covered by CPs *P LeadsTo Q ExactlyAfter k* (El Gammal 2012) or *An activity of type  $a_1$  must be started/completed before/after/on  $t$  time units (relative to  $t_0$ )* (Caron et al. 2013a).

#### 4 Data-related CP Design

The CP design step takes the results of the constraint elicitation step as input, i.e., the atomic data-oriented constraints from Table 2. In Sect. 4.1 the atomic data-oriented constraints are mapped onto existing CPs in order to determine which of them are not covered by CPs yet (*Artifact 6: list of CP gaps*). All of those atomic data-oriented constraints on the gap list are then designed as new data-related CPs in Sect. 4.2 (*Artifact 7*).

##### 4.1 Mapping of Data-Oriented Constraints to Existing CPs

Before the mapping of atomic data-oriented constraints to existing data-related CPs, constraints with the same semantics get deleted. Accordingly, constraint 13 *Data must be unique over time* just extends the existing uniqueness CPs by including a time component, which in turn is superfluous because uniqueness has to be given at any point in time. *Over time* means in the context of this paper a given time span. Therefore, we subsume this atomic data-oriented constraint under the constraint 11 *Data must be unique*. Also the constraints 18 *Data origin must be known* and 19 *Data purpose must be known* are subcategories of the constraint 3 *Data must be available/complete*, because both, origin and purpose of data can be specified as an own CP, but can be simply seen as existence of data representing either the purpose or the origin which have to be available. Nevertheless, we do not merge those 3 atomic data-oriented constraints due to simplicity of applicability by business users. They do not need to know what constraints have to be merged and have the same semantics—that is done by the CPs. In the end, 19 atomic data-oriented constraints are extracted.

**Table 3** Atomic data-oriented constraints mapped to CPs

No.	Atomic data-oriented constraint	CP name
1	Data must be accurate/consistent/must have integrity	Only some examples are listed due to various matching CPs (“Appendix E” includes the full list): Event data multiplicity rule; Logical derivation rule; Event data equality rule
3	Data must be available/complete	Mandatory event data rule
4	Data must be calculated using formula	Arithmetic derivation rule
5	Data must be equal/have the same values	Event data equality rule
7	Data must be in domain	Event data value set rule
8	Data must be in range	Event data value range rule
9	Data must be missing	¬ Mandatory event data rule
12	Data must be unique	Event data uniqueness rule
13	Data must be unique over time	Event data uniqueness rule
14	Data must conform to a specific format	Event data format rule
15	Data must not be changed	Time-oriented integrity rule & Activity-oriented integrity rule
16	Data must not be used	Data element never read rule
18	Data origin must be known	Mandatory event data rule
19	Data purpose must be known	Mandatory event data rule

The mapping between the atomic data-oriented constraints from Table 2 found in various regulatory documents (e.g., AnaCredit, BCBS 239, GTelG) and the existing CPs from literature is based on the associated CP descriptions. Table 3 summarizes the results where for 14 out of 19 data-oriented constraints a mapping can be found. The remaining 5 data-oriented constraints form the list of CP gaps (*Artifact 6*), i.e., 2, 6, 10, 11, and 17.

Based on the mapping between atomic data-oriented constraints to existing CPs, research question *Q1: Is there a gap between coverage of business process perspectives in literature and demands from regulatory documents? Particularly for data-oriented compliance demands?* can be answered with *yes*. *There are gaps for the data perspective of business processes. The other business process perspectives seem to be covered.*

For research question *Q2: Which data-oriented compliance constraints in regulatory documents are not covered by existing CPs?* the answer is: *there are 5 data-oriented constraints identified as gaps and put on the gap list (Artifact 6). Those are not covered by existing CPs for the data perspective of business processes.*

The next question is which of the remaining 5 constraints result in the creation of a data-related CP. Therefore all of them are investigated in detail by applying the following CP design criteria.

1. *Occurrence in regulatory documents* – the atomic data-oriented constraints must be stated in at least two regulatory documents. Doing so, the property of

repetition and reusability of a pattern in general will be ensured.

2. *Usage in application domains* – the atomic data-oriented constraints must at least be applied in two different domains/industries. This underlines the importance of the constraints for multiple domains.
3. *Absolute occurrences* – the atomic data-oriented constraints must occur at least three times over all documents. Thus, repetition of the data-oriented constraint is ensured also within a regulatory document, besides the usage in at least two different domains.

If an atomic data-oriented constraint fulfils all three criteria, it is designed as CP. Constraints not fulfilling all criteria are not further considered and could be subject for future research. However, all identified atomic data-oriented constraints fulfil all three criteria and are designed as CPs in the next Sect. 4.2.

#### 4.2 Data-related CPs

For each CP its name, description, and related CPs are described ( $\mapsto$  *Artifact 7*). The newly designed CPs are formalized using EC – a well-known logical language. We decided to use EC based on the findings of Fdhila et al. (2016) which show the suitability of EC to specify (instance-spanning) constraints referring to data.

Table 4 shows the design of two new CPs for constraint 2 that is split into two CPs, i.e., *Data must be available at a certain time* and *Data must be available for a certain time*, where the latter one can be substituted by a sequence of the

**Table 4** Design of CP for atomic data-oriented constraint 2

Atomic data-oriented constraint no.	2
Atomic data-oriented constraint	Data must be available <b>at</b> /for a certain time
CP name	Mandatory time-oriented event data CP
CP description	The value of event data type $p_1$ (wrt an event of type $e_1$ for an activity of type $a_1$ ) must exist at time $t_R$
Related existing CPs	<i>Mandatory event data rule; Time-oriented integrity rule; Missing Data</i>
Formal specification	EVENTS: ReadData(value $_{p_1}$ ) FLUENTS: IsAvailable(value $_{p_1}$ ) STATEMENTS: Happens(ReadData(value $_{p_1}$ ), $t_R$ ) $\rightarrow$ HoldsAt(IsAvailable(value $_{p_1}$ ), $t_R$ )
Atomic data-oriented constraint no.	2
Atomic data-oriented constraint	Data must be available <b>at/for</b> a certain time
CP name	Mandatory time span event data CP
CP description	The value of event data type $p_1$ (wrt an event of type $e_1$ for an activity of type $a_1$ ) must exist for $t$ time units (relative to $t_0$ )
Related existing CPs	<i>Mandatory event data rule; Time-oriented integrity rule; Missing Data; Not Deleted on Time</i>
Formal specification	EVENTS: ReadData(value $_{p_1}$ ) FLUENTS: IsAvailable(value $_{p_1}$ ) STATEMENTS: $\forall t_R$ Happens(ReadData(value $_{p_1}$ ), $t_R$ ) $\wedge t_0 \leq t_R \leq t_0 + t \rightarrow$ HoldsAt(IsAvailable(value $_{p_1}$ ), $t_R$ )

first one. For atomic data-oriented constraint 2 we check if an activity to read data resulting in an event *ReadData* can be successfully executed at time point  $t_R$  and for the second type of constraint 2 the read event has to occur in a given time span  $ts$  starting at  $ts_{start}$  and ending at  $ts_{end}$ , respectively.

“Appendix G” lists the CP designs of the other atomic data-oriented constraints 6, 10, 11 and 17. It includes also the different characteristics of a CP as shown in Table 4. The answer to research question *Q3: How can missing CPs be defined for uncovered data-oriented compliance constraints?* is: *like for all other business process perspectives through application of a uniformed elicitation method and use of a suitable specification formalism.*

## 5 Discussion

We identify and discuss the following threats to validity of our study:

*Search terms and engine* The literature search is based on search strings defined by following the main concepts of the approach introduced by Kitchenham and Charters (2007). For this, common terms in the area of compliance management for business processes were identified and combined to the search strings. However, a systematic

analysis of potential search terms was not conducted, and this may have led to a different set of search terms influencing the results of the literature review. Especially conformance checking was not considered in the search terms, since it “is typically understood as the problem of comparing an existing process model with an event log” (Ly et al. 2015). The goal is to determine the differences between those two and measure the degree of deviation (Ly et al. 2015). Hence, the focus of this paper is not on identifying differences in event logs and process models, but to obtain an overview of currently existing CPs with a special focus on the data perspective of business processes. Also security was not considered as search term. There exists literature about the usage of patterns in secure system development. However, those approaches often focus on the software development process and use patterns more in the sense of design patterns instead of CPs (e.g., Ahmed and Matulevičius 2014; Matulevičius 2017). Other literature uses so called security policy types (e.g., Salnitri and Giorgini 2014), that can be transformed into CPs and open another area of future research. Furthermore, we see security as an application area for CPs to deal with security concerns/issues. *Data* is another important search term which is included in the search strings to highlight the data perspective. This term is dedicatedly used to retrieve search hits related to the data perspective in business processes.

The creation of a common term set/dictionary for the BPC management domain (see differences/overlappings of compliance requirement, compliance constraint, CP, compliance rule) can overcome this problem and may lead to a better understanding by all involved parties. Google Scholar was used as primary search engine. Involving also other search engines, archives, and online libraries may have led to a larger search space and thus more search hits. Nevertheless, we assume that most of the publishers of journals, magazines, and conference proceedings bear in mind search engine optimization and thus their documents will be detectable via a Google Scholar search. Finally, 18 papers which are included in the search hits were not accessible for us, i.e., they were neither accessible via Google Scholar nor via *u:search*. Those papers may include additional CPs because 16 of them entered Stage 2 of our literature review process. However, these are only two percent of all search hits and the chance of additional information is quite low.

*Literature investigation* We perform a literature investigation with focus on CP, whereas Fellmann and Zasada (2014) followed a more general approach. They tried to obtain an overview of existing BPC approaches regardless if CPs are used or not. However, they also grouped the various approaches according to dimensions like scope (i.e., process modelling patterns like order, time, and resource), life cycle phases, and formality (i.e., highly formalized or management-oriented approaches).

The *selection of regulatory documents* which are used for the elicitation of data-oriented constraints focuses on regulations that are mainly effective in Austria and the financial industry due to accessibility, up-to-dateness, and our domain knowledge. As a starting point 10 regulatory documents were selected, but even more should be investigated in future research. There are many other regulatory documents available that may also include a wide range of data-oriented constraints. That will increase the probability that the identified data-related CPs will be effective for further documents. It can also be possible that further data-related CPs can be identified based on further documents.

*Scope of CP collection* The present paper aims at the identification of existing CPs and gaps due to regulatory demands. Therefore a comprehensive CP collection is created and compared to constraints from 10 selected regulatory documents. We do not claim completeness of the CP collection, since regulatory documents are vivid documents and change over time which makes it necessary to create, adapt, or delete CPs. Therefore, the underlying scope of the CP collection should be continuously extended and adapted by analyzing further regulatory documents from more domains/industries.

*Process perspectives* Literature uses different perspectives to categorize CPs. These perspectives mainly relate to

the well known business process perspectives control flow, resources, time, and data (Fig. 4). We decided to follow this categorization approach and regard the additionally identified properties like atomic, composite, and anti-pattern as a more fine-grained subcategorization. Also the property of instance-spanning constraints (cf. Fdhila et al. 2016; Rinderle-Ma et al. 2016) is deliberately omitted to reduce complexity. Of course other CP hierarchies which interchange category and subcategory or use other hierarchy structures exist (e.g., Awad et al. 2015; Barnawi et al. 2016; El Gammal 2012; Turetken et al. 2011), but do not change the intended focus of a CP on a certain business process aspect. At *bpmpatterns.org* a pattern taxonomy for business process model patterns is described. The taxonomy consists of 9 categories (e.g., structure and behavior, resource patterns, data patterns, content patterns, integration and conversion patterns) which are used to group the patterns. The listed categories overlap with the perspectives and properties we used, however the taxonomy has a broader focus on patterns for business processes in general. In contrast, our paper specifically investigates CPs. Nevertheless, the presented categories of the taxonomy can be considered in future research in order to extend the perspectives and properties, and allow for a more fine-granular and/or multiple assignment of perspectives and properties to CPs.

*Matching* Another crucial step is the matching of CPs with the same meaning. This mapping is done mainly using the provided CP names and descriptions. Also formal specifications are used to find matches, but those are rare. Thus, we perform that task to the best of our knowledge. To always find the correct matching CPs, a systematic analysis and matching must be executed based on formal specifications like LTL, IP and ECL. However, sometimes the exact meaning of a described CP cannot be determined and we excluded those CPs from our approach. In total, 7 CPs remain unmapped due to insufficient specification/description of the CP and for 49 CPs a mapping is not applicable. A full list of all unmapped CPs is shown in “Appendix F”. Future research may also consider those CPs for an even more comprehensive coverage of collection and categorization.

*CP criteria and design* The selection of CP design criteria just focused on basic concepts of usage and occurrence of underlying atomic data-oriented constraints. More sophisticated criteria can be established for CP design. Nevertheless, the three defined criteria are sufficient for the purpose of this paper. The design of the CPs itself tries to follow the approach presented in Caron et al. (2013a). Therefore, the same naming conventions and structuring of CPs are used. However, there is no common concept in literature how to define CPs. At the first sight the derived atomic data-oriented constraints may look like simple



integrity checks for databases. Since constraints are defined in a generic manner, they can also serve as basis for integrity check rules executed on database level. Nevertheless, these constraints are applied to business processes themselves, because business processes use data as input, create data in process instances, and/or produce data as output of a certain activity or the entire process instance. Thus, it is a point of view where those checks for data items take place. There is no clear separation between compliance checks that must be performed inside and outside a business process. Often data-oriented constraints are outsourced to more specialized tools/systems (e.g., database management systems, extract-transform-load tools). However, regulatory documents force organizations to intensify their efforts in compliance management concerning data aspects, and approaches exist to propagate process constraints from the process level to the database level (cf. Gómez-López et al. 2015).

*Relation to DQ approaches* Various DQ approaches categorize DQ according to different dimensions like accuracy, timeliness, or completeness (Lee et al. 2002; Wang 1998; Bai et al. 2018; Fox et al. 2018). Those dimensions highlight various aspects of data and the measurement of DQ is conducted in accordance with the dimensions. Usually DQ is measured on data stored in databases and ensured by dedicated DQ management tools. However, the application of data-related CPs to business processes can contribute to high quality data by ensuring DQ before data storage. Furthermore, data-related CPs could be included as an integral part of existing DQ frameworks besides already established DQ controls and measures (at database level).

## 6 Conclusion and Outlook

In this paper we present an overview and status of current research in the field of compliance management using a pattern-based approach. A comprehensive literature review is conducted and various information about CPs are extracted from the selected literature. All identified CPs are organized into a single collection embracing various perspectives and properties to allow a categorization along the main perspectives of business processes. The literature analysis shows differences in the treatment of the process perspectives control flow, resources, time, and data. Especially the underrepresentation of the data perspective motivates a further investigation into regulatory documents to analyze the relevance of data-related CPs. In total 10 different regulatory documents covering multiple domains serve as an initial basis for the elicitation of data-oriented constraints with the goal of identifying a gap in existing CPs. These atomic data-oriented constraints are mapped

onto existing CPs and all mismatches are collected in a gap list. Afterwards new data-related CPs are designed based on this gap list to support the enforcement of the original constraints in business processes. Since various current and even more future regulatory documents (will) put the data-oriented constraints in the spotlight, the closure of these gaps is a necessary step to ensure organizations' compliance with laws, guidelines, and standards. Future research should focus even more on the data aspects of compliance constraints. Especially, research regarding the relationship between CPs should be conducted to understand the interaction of CPs and its influence on root-cause analysis as well as the support for root-cause mitigation and remediation.

**Acknowledgements** Open access funding provided by University of Vienna.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Ahmed N, Matulevičius R (2014) Securing business processes using security risk-oriented patterns. *Comput Stand Interfaces* 36(4):723–733
- AnaCredit. Regulation (EU) 2016/867 of the European Central Bank of 18 May 2016 on the collection of granular credit and credit risk data (ECB/2016/13). [https://www.ecb.europa.eu/ecb/legal/pdf/celex\\_32016r0867\\_en\\_txt.pdf](https://www.ecb.europa.eu/ecb/legal/pdf/celex_32016r0867_en_txt.pdf). Accessed 13 Dec 2017
- Awad A (2010) A compliance management framework for business process models. PhD thesis, University of Potsdam
- Awad A, Weske M (2010) Visualization of compliance violation in business process models. In: BPM 2009 international workshops on business process management workshops, Ulm, Germany, September 7, 2009. Revised Papers. Springer, Heidelberg, pp 182–193
- Awad A, Smirnov S, Weske M (2009) Resolution of compliance violation in business process models: A planning-based approach. In: Proceedings OTM 2009: confederated international conferences on the move to meaningful internet systems, CoopIS, DOA, IS, and ODBASE 2009, Vilamoura, Part I, pp 6–23
- Awad A, Weidlich M, Weske M (2011) Visually specifying compliance rules and explaining their violations for business processes. *J Vis Lang Comput* 22(1):30–55
- Awad A, Barnawi A, Elgammal A, Elshawi R, Almalaise A, Sakr S (2015) Runtime detection of business process compliance violations: an approach based on anti patterns. In: Proceedings of the 30th annual ACM symposium on applied computing, SAC '15, Salamanca, pp 1203–1210
- Bai L, Meredith R, Burstein F (2018) A data quality framework, method and tools for managing data quality in a health care setting: an action case study. *J Decis Syst* 27(sup1):144–154

- Bank for International Settlements (2013) Principles for effective risk data aggregation and risk reporting. <http://www.bis.org/bcbs/publ/d399.pdf>. Accessed 31 Jul 2017
- Barnawi A, Awad A, Elgammal A, Elshawi R, Almalaise A, Sakr S (2016) An anti-pattern-based runtime business process compliance monitoring framework. *Int J Adv Comput Sci Appl* 7:551–572
- Becker J, Ahrendt C, Coners A, Weiß B, Winkelmann A (2016) Business rule based extension of a semantic process modeling language for managing business process compliance in the financial sector. In: Fhnrich K-P, Franczyk B, (eds) *INFORMATIK 2010. Service Science Neue Perspektiven fr die Informatik. Band 1*, Bonn. Gesellschaft fr Informatik e.V., pp 201–206
- Becker J, Delfmann P, Eggert M, Schwittay S (2012) Generalizability and applicability of model-based business process compliance-checking approaches: a state-of-the-art analysis and research roadmap. *Bus Res* 5:221–247
- Bernardi ML, Cimitile M, Di Francescomarino C, Maggi FM (2014) Using discriminative rule mining to discover declarative process models with non-atomic activities. In: *Proceedings 8th international symposium, RuleML 2014, Co-located with the 21st European conference on artificial intelligence rules on the web. From theory to applications, ECAI 2014, Prague*, pp 281–295
- BSI Act (2009) BSI Act of 14 August 2009 (Federal Law Gazette I p. 2821) last amended by Article 1 of the Act of 23 June 2017 (Federal Law Gazette I p. 1885). [https://www.gesetze-im-internet.de/bsig\\_2009/BJNR282110009.html](https://www.gesetze-im-internet.de/bsig_2009/BJNR282110009.html). Accessed 21 Jul 2017
- Cabanillas C, Resinas M, Ruiz-Corts A (2010) On the identification of data-related compliance problems in business processes. In: *Conference: VI Jornadas Científico-Técnicas en Servicios Web y SOA (JSWEB'10)*, Valencia, 01
- Caron F, Vanthienen J, Baesens B (2013a) Comprehensive rule-based compliance checking and risk management with process mining. *Decis Support Syst* 54(3):1357–1369
- Caron F, Vanthienen J, Baesens B (2013b) Advances in rule-based process mining: applications for enterprise risk management and auditing. In: *KU Leuven: Faculty of Economics and Business Working Paper No. KBI 1305*
- Cheikhrouhou S, Kallel S, Guermouche N, Jmaiel M (2014) Enhancing formal specification and verification of temporal constraints in business processes, anchorage. In: *2014 IEEE international conference on services computing*, pp 701–708
- Chesani F, Mello P, Montali M, Riguzzi F, Sebastianis M, Storari S (2008) Compliance checking of execution traces to business rules: an approach based on logic programming. In: *Proceedings of the Sixth international conference on business process management*, Perugia
- Chesani F, Mello P, Montali M, Riguzzi F, Sebastianis M, Storari S (2009) Checking compliance of execution traces to business rules. In: *Business process management workshops: BPM 2008 international workshops. Milano, Revised Papers*, pp 134–145
- De Masellis R, Maggi FM, Montali M (2014) Monitoring data-aware business constraints with finite state automata. In: *Proceedings of the 2014 international conference on software and system process, Nanjing, ICSSP 2014. ACM*, pp 134–143
- DPA (2000) Bundesgesetz über den Schutz personenbezogener Daten (Datenschutzgesetz 2000 - DSG 2000), Fassung vom. <https://www.ris.bka.gv.at/Bundesrecht/>. Accessed 13 Jul 2017
- Dwyer MB, Avrunin GS, Corbett JC (1998) Property specification patterns for finite-state verification. In: *Proceedings of the second workshop on formal methods in software practice, Clearwater Beach, FMSP '98*, pp 7–15
- E-GovG (2017) Bundesgesetz über Regelungen zur Erleichterung des elektronischen Verkehrs mit öffentlichen Stellen (E-Government-Gesetz - E-GovG), Fassung vom. <https://www.ris.bka.gv.at/Bundesrecht/>. Accessed 13 Dec 2017
- El Gammal AFSA (2012) Towards a comprehensive framework for business process compliance. PhD thesis, Tilburg University, School of Economics and Management
- ELGA-VO 2015. Verordnung der Bundesministerin für Gesundheit zur Implementierung und Weiterentwicklung von ELGA (ELGA-Verordnung 2015 ELGA-VO 2015), Fassung vom. <https://www.ris.bka.gv.at/Bundesrecht/>. Accessed 13 Dec 2017
- Elgammal A, Turetken O, van den Heuvel W-J, Papazoglou M (2010) Root-cause analysis of design-time compliance violations on the basis of property patterns. In: *Proceedings 8th international conference, ICSSOC 2010 service-oriented computing, San Francisco. Springer, Heidelberg*, pp 17–31
- Elgammal A, Turetken O, van den Heuvel W-J, Papazoglou M (2016) Formalizing and applying compliance patterns for business process compliance. *Softw Syst Model* 15(1):119–146
- Fdhila W, Gall M, Rinderle-Ma S, Mangler J, Indiono C (2016) Classification and formalization of instance-spanning constraints in process-driven applications. In: *Proceedings business process management: 14th international conference, BPM 2016. Springer, Cham*
- Fellmann M, Zasada A (2014) State-of-the-art of business process compliance approaches. In: *Proceedings of the European conference on information systems (ECIS) 2014, Tel Aviv, AISeL*
- Fox F, Aggarwal VR, Whelton H, Johnson O (June 2018) A data quality framework for process mining of electronic health record data. In: *2018 IEEE international conference on healthcare informatics (ICHI)*, New York, pp 12–21
- Giblin C, Müller S, Pfitzmann B (2006) From regulatory policies to event monitoring rules: towards model-driven compliance automation. In: *IBM Research Zurich, Report RZ*, pp 3662
- Gomez-Lopez MT, Gasca RM, Rinderle-Ma S (2013) Explaining the incorrect temporal events during business process monitoring by means of compliance rules and model-based diagnosis. In: *2013 17th IEEE international enterprise distributed object computing conference workshops, Vancouver*, pp 163–172
- Gómez-López MT, Gasca RM, Pérez-Álvarez JM (2015) Compliance validation and diagnosis of business data constraints in business processes at runtime. *Inf Syst* 48:26–43
- Gong P, Knuplesch D, Reichert M (2016) Rule-based monitoring framework for business process compliance. In: *Technical Report UIB-2016-3, Ulm University*
- Gruhn V, Laue R (2005) Specification patterns for time-related properties. In: *12th International symposium on temporal representation and reasoning (TIME'05)*. Bington, pp 189–191
- Gruhn V, Laue R (2006) Patterns for timed property specifications. *Electron Notes Theor Comput Sci* 153(2):117–133
- GTelG (2012) Bundesgesetz betreffend Datensicherheitsmaßnahmen bei der Verwendung elektronischer Gesundheitsdaten (Gesundheitstelematikgesetz 2012 GTelG 2012), Fassung vom. <https://www.ris.bka.gv.at/Bundesrecht/>. Accessed 13 Dec 2017
- Hashmi M, Governatori G, Wynn MT (2012) Business process data compliance. In: *Rules on the Web: research and applications. Springer, Heidelberg*, pp 32–46
- IMA-VO (2011) Verordnung der E-Control, mit der die Anforderungen an intelligente Messgeräte bestimmt werden (Intelligente Messgeräte-AnforderungsVO 2011 IMA-VO 2011), Fassung vom. <https://www.ris.bka.gv.at/Bundesrecht/>. Accessed 13 Dec 2017
- Khan A, Ehsan N, Mirza E, Sarwar SZ (2012) Integration between customer relationship management (CRM) and data warehousing. *Procedia Technol* 1:239–249
- Kitchenham B, Charters S (2007) Guidelines for performing systematic literature reviews in software engineering. In: *Technical*

- report 2007-01, Keele University, School of Computer Science and Mathematics
- Knuplesch D, Reichert M (2017) A visual language for modeling multiple perspectives of business process compliance rules. *Soft Syst Model* 16(3):715–736
- Kumar A, Barton RR (2017) Controlled violation of temporal process constraints models, algorithms and results. *Inf Syst* 64:410–424
- Kumar A, Yao W, Chu C-H, Li Z (2010) Ensuring compliance with semantic constraints in process adaptation with rule-based event processing. In: *Proceedings international symposium, RuleML 2010 semantic web rules*, Washington, DC, pp 50–65
- Kumar A, Sabbella SR, Barton RR (2015) Managing controlled violation of temporal process constraints. In: *Proceedings 13th international conference business process management, BPM 2015*, Innsbruck, pp 280–296
- Lam VS (2017) Detecting violation of business constraints in declarative process execution: a case study. *Bus Process Manag J* 23(5):1043–1063
- Lee YW, Strong DM, Kahn BK, Wang RY (2002) AIMQ: a methodology for information quality assessment. *Inf Manag* 40(2):133–146
- Lu R, Sadiq S, Governatori G (2009) On managing business processes variants. *Data Knowl Eng* 68(7):642–664
- Ly LT (2016) *SeaFlows: a compliance checking framework for supporting the process lifecycle*. PhD thesis, University of Ulm
- Ly LT, Rinderle-Ma S, Dadam P (2010) Design and verification of instantiable compliance rule graphs in process-aware information systems. In: *Proceedings 22nd international conference advanced information systems engineering, CAiSE 2010*, Hammamet, pp 9–23
- Ly LT, Knuplesch D, Rinderle-Ma S, Göser K, Pfeifer H, Reichert M, Dadam P (2011) *Seaflows toolset: compliance verification made easy for process-aware information systems*. In: *Information systems evolution: CAiSE Forum 2010*. Hammamet, Selected Extended Papers, pp 76–91
- Ly LT, Maggi FM, Montali M, Rinderle-Ma S, van der Aalst WM (2015) Compliance monitoring in business processes: functionalities, application, and tool-support. *Inf Syst* 54:209–234
- Maggi FM, Montali M, Westergaard M, van der Aalst WMP (2011) Monitoring business constraints with linear temporal logic: an approach based on colored automata. In: *Proceedings 9th international conference, BPM 2011 business process management*, Clermont-Ferrand. Springer, Heidelberg, pp 132–147
- Marín-Ortega PM, Dmitriyev V, Abilov M, Gómez JM (2014) ELTA: new approach in designing business intelligence solutions in era of big data. *Procedia Technol* 16:667–674
- Matulevičius R (2017) *Secure system development*. Springer, Berlin
- Montali M, Maggi FM, Chesani F, Mello P, Aalst WMPVD (2014) Monitoring business constraints with the event calculus. *ACM Trans Intell Syst Technol* 5(1):17:1–17:30
- Namiri K, Stojanovic N (2007) Pattern-based design and validation of business process compliance. In: *Proceedings OTM confederated international conferences CoopIS on the move to meaningful internet systems 2007, CoopIS, DOA, ODBASE, GADA, and IS, DOA, ODBASE, GADA, and IS 2007*, Vilamoura, Part I. Springer, Heidelberg, pp 59–76
- Oesterreichs Energie (2015) Smart metering use-cases für das advanced meter communication system (AMCS). [https://oesterreichsenergie.at/files/Downloads%20Netze/Oesterreich%20Use%20Cases%20Smart%20Metering\\_14122015\\_Version\\_1-1.pdf](https://oesterreichsenergie.at/files/Downloads%20Netze/Oesterreich%20Use%20Cases%20Smart%20Metering_14122015_Version_1-1.pdf). Accessed 21 Jul 2017
- Oesterreichs Energie (2018) Requirements catalog end-to-end security for smart metering. [https://oesterreichsenergie.at/files/Downloads%20Netze/E2E-Sicherheit-Anforderungskatalog-EN\\_1.1\\_final.pdf](https://oesterreichsenergie.at/files/Downloads%20Netze/E2E-Sicherheit-Anforderungskatalog-EN_1.1_final.pdf). Accessed 21 Jul 2017
- Papazoglou MP (2011) Making business processes compliant to standards and regulations. In: *2011 IEEE 15th international enterprise distributed object computing conference*, Helsinki, pp 3–13
- Paulson LD (2000) Data quality: a rising e-business concern. *IT Prof* 2(4):10–14
- Ramezani E (2017) *Understanding non-compliance*. PhD thesis, Eindhoven University of Technology
- Rinderle-Ma S, Kabicher-Fuchs S (2016) An indexing technique for compliance checking and maintenance in large process and rule repositories. *Enterp Model Inf Syst Arch* 11(1):1–24. <https://doi.org/10.18417/emisa.11.2>
- Rinderle-Ma S, Gall M, Fdhila W, Mangler J, Indiono C (2016) Collecting examples for instance-spanning constraints. In: *Technical report*. [arXiv.org. http://eprints.cs.univie.ac.at/4634/](http://eprints.cs.univie.ac.at/4634/)
- Salnitri M, Giorgini P (2014) Transforming socio-technical security requirements in secBPMN security policies. In: *CEUR workshop proceedings*, vol 1157
- Santos EAP, Francisco R, Vieira AD, de FR Loures E, Buseti MA, (2012) Modeling business rules for supervisory control of process-aware information systems. In: *BPM 2011 international workshops, Clermont-Ferrand, revised selected papers, part II business process management workshops*. Springer, Heidelberg, pp 447–458
- Schleicher D, Leymann F, Schumm D, Weidmann M (2010) Compliance scopes: extending the bpmn 2.0 meta model to specify compliance requirements. In: *International conference on service-oriented computing and applications*, Perth, pp 1–8
- Schumm D, Turetken O, Kokash N, Elgammal A, Leymann F, van den Heuvel WJ (2010) Business process compliance through reusable units of compliant processes. In: Daniel F, Facca FM (eds) *Current trends in web engineering*. Lecture notes in computer science, vol 6385. Springer, Heidelberg, pp 325–337. [https://doi.org/10.1007/978-3-642-16985-4\\_29](https://doi.org/10.1007/978-3-642-16985-4_29)
- Stausberg J, Pritzkeleit R, Schmidt CO, Schrader T, Nonnemacher M (2011) Indicators of data quality: revision of a guideline for networked medical research. *Stud Health Technol Inform* 180:711–715
- Stuht T, Speck A, Feja S, Witt S, Pulvermüller E (2012) Rule determination and process verification using business capabilities. In: *Proceedings 5th IFIP WG 8.1 working conference the practice of enterprise modeling, PoEM 2012*, Rostock. Springer, Heidelberg, pp 46–60
- Thullner R, Rozsnyai S, Schiefer J, Obwegger H, Suntinger M (2011) Proactive business process compliance monitoring with event-based systems. In: *2011 IEEE 15th international enterprise distributed object computing conference workshops*, Helsinki, pp 429–437
- Trčka N, van der Aalst WMP, Sidorova N (2009) Data-flow anti-patterns: discovering data-flow errors in workflows. In: *Proceedings 21st international conference advanced information systems engineering, CAiSE 2009*, Amsterdam. Springer, Heidelberg, pp 425–439
- Turetken O, Elgammal A, van den Heuvel W-J, Papazoglou MP (2011) Enforcing compliance on business processes through the use of patterns. In: *Proceedings ECIS*, vol 5. <http://aisel.aisnet.org/ecis2011/5>
- Turetken O, Elgammal A, van den Heuvel W-J, Papazoglou MP (2012) Capturing compliance requirements: a pattern-based approach. *IEEE Softw* 29(3):28–36
- van der Aalst W, Artale A, Montali M, Tritini S (2017) Object-centric behavioral constraints: integrating data and declarative process modelling. In: *Proceedings of the 30th international workshop on description logics*, Montpellier

Wang RY (1998) A product perspective on total data quality management. *Commun ACM* 41(2):58–65

Yu J, Manh TP, Han J, Jin Y, Han Y, Wang J (2006) Pattern based property specification and verification for service composition.

In: Proceedings 7th international conference on web information systems engineering web information systems - WISE 2006, Wuhan. Springer, Heidelberg, pp 156–168